Department of Informatics, King's College London
Pattern Recognition (6CCS3PRE/7CCSMPNN).
Assignment: Support Vector Machines (SVMs)

The Iris flower data set consists the date of 3 species of Iris (setosa, virginica and versicolor) characterised by 4 features: the length and the width of the sepals and petals in centimetres. The Iris data set (in Matlab format) with file name "`iris_class1_2_3_4D.mat`" is available on KEATS. A multi-class SVM-based classifier formed by multiple SVMs is employed to handle the classification problem.

Two 3^{rd}-party Matlab SVM toolboxes are recommended to implement the multi-class SVM-based classifier. You can use either one of them for this assignment. Details can be found in Appendix 1.

a) Write down your 7-digit student ID denoted as $s_1s_2s_3s_4s_5s_6s_7$.

b) Find R_1 which is the remainder of $\frac{s_1+s_2+s_3+s_4+s_5+s_6+s_7}{4}$. Table 1 shows the multi-class methods to be used corresponding to the value of R_1 obtained.

R_1	Method
0	One against one
1	One against all
2	Binary decision tree
3	Binary coded

Table 1: R_1 and its corresponding multi-class method.

c) Write a Matlab code to randomly partition the Iris data into 5 equal size sub-data sets D_1 to D_5, which will be used for 5-fold cross validation. *Hint: A built-in Matlab function* `randperm` *can help generate random indices for data partitioning. Type "*`help randperm`*" for the syntax and details.*

d) Write a Matlab script to implement the multi-class SVM-based classifier using the R_1^{th} method. Train and test multi-class SVM-based classifier using 5-fold cross validation. The cross-validation process will repeat the training-testing process 5 times. In the i^{th} time, D_i sub-data set is used for testing and the rest sub-data sets (forming a single data set) are be used for training. For example, in the 1^{st} time, D_1 is used for testing and D_2, D_3, D_4 and D_5 forming a single data set is used for training.

e) Summarise the classification accuracy (in %) in the form of Table 2 for different kernels (linear, polynomial and RBF kernels) and values of C. Comment on the results.

- Consider the values of C as 1, 10, and 100.
- Pick the coefficients of polynomial and RBF kernels of your choice. Use the same coefficients for all values of C.

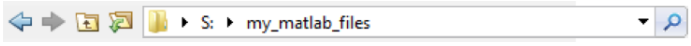
In total, you will have 3 tables in the form of Table 2. Each for a value of C.

	Linear Kernel		Raaa (k)		Raaa (k)	
Fold	Training Accuracy (%)	Testing Accuracy (%)	Training Accuracy (%)	Testing Accuracy (%)	Training Accuracy (%)	Testing Accuracy (%)
1						
2						
3						
4						
5						
Average Accuracy						

Table 2: Summary accuracy table for linear kernel, polynomial kernel and RBF kernel with a chosen value of C.

Appendix 1

LibSVM:

LibSVM is recommended that if you use **64-bit Windows machines**. It can be downloaded from `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`. First unzip the downloaded zip file and you will find a folder "`windows`". Point the current working folder to "`windows`" in the Matlab command windows using the "current folder toolbar" which looks like 

. All your Matlab files should be place in the folder "`windows`". An example Matlab script using LibSVM can be found on KEATS.

OSU-SVM

OSU-SVM is recommended that if you use **32-bit Windows machines**, for example, the Matlab in Global Desktop (`https://desktop.kcl.ac.uk/vpn/index.html`). It can be downloaded from `http://svm.sourceforge.net/download.shtml`. First unzip the downloaded zip file and point the current working folder to the unzipped OSU-SVM folder in the Matlab command windows using the "current folder toolbar". All your Matlab files should be place in the unzipped OSU-SVM folder. An example Matlab script using OSU-SVM can be found on KEATS.